The Birth Number Concept and Record Linkage

ANDERS S. LUNDE, PhD

Government-issued birth numbers and the development of record linkage systems in the United States and other countries are examined.

Introduction

The practice of linking separate records on individuals has been growing in this country and abroad since World War II. Stimulated by the expansion of computer capability, this development, rather well advanced in the fields of business and government, has engaged the interest of health researchers. While a considerable amount of thinking has gone into constructing closed linkage systems which relate records within a single institution or agency, there are those who conceive of open universal systems in which records collected by one agency in one part of the country may be linked according to source and purpose to entirely different records from another part of the country. The objectives to be achieved by these systems differ widely, from supplying a list of drugs previously issued to a patient to providing information on related health episodes from birth to death. Beyond that the broader record linkage systems have been described as laying the basis for a national death clearance index, making possible a comprehensive study of the natural history of disease, providing details for family-related genetic studies, and generating data for epidemiological research, to mention a few of the possibilities.

Although Charles Babbage (1792–1871) and others developed the ideas underlying modern computer theory, it was William Farr (1803–1887) who first noted that organizing and combining data on important health events of individuals in the whole population would permit cross-sectional and longitudinal studies of unlimited scope. In the 19th century no practical way for combining records existed, and while there were great advances in data collecting and handling little could be done in the area of mass linkage. Accounting systems combining data by machine came into being in the 20th century but such systems had mechanical and cost limitations. The advent of the computer made mass linkage feasible for the first time. Not only did it become possible to store detailed data but also to link diverse sources of information, in the form of reports, records, certificates, charts, and any other document pertaining to an individual.

Beginning in the 1940s a number of health record linkage systems came into existence. The primary identification in these systems differed. In some cases the first link was based on the letters of family names; in others the primary link was a code number such as a patient record or chart number. A large number of codes were invented. Many were sufficient to serve local needs but could not practically be extended. Others were discarded because they were unmanageable, time-consuming, and costly. Nevertheless, the feasibility of extended record linkage systems was proven without a doubt. In Canada, a national index combining birth, marriage, and death certificates with a name link for alphabetical retrieval, established in connection with the family allowance system, was computerized. Later, in a project of Atomic Energy of Canada to study the effects of radiation, Newcombe and other researchers showed that a computer program could be written to link stillbirth, live birth, and marriage records for the entire population of British Columbia.[1] The Oxford Linkage Study also supported this development.[2]

In extensive systems combining items from a number of records, a number code as a primary referrant (the standard

At the time of this study, Dr. Lunde was Director of the Office of State Services, National Center for Health Statistics, Research Triangle Park, NC 27709. He is now Acting Director of the Office of International Statistics, National Center for Health Statistics, Health Resources Administration, Department of Health, Education, and Welfare, Rockville, MD 20852. The opinions expressed are those of the author and do not necessarily reflect the views of the National Center for Health Statistics and the Department of Health, Education, and Welfare. This article was presented at the Statistics Section Session of the 101st Annual Meeting of the American Public Health Association, in San Francisco, California, November 8, 1973.

identifier) has been used most extensively. Other codes from a name or reference number serve as backup identifiers. The ideal standard identifier has specified characteristics. It must be unique, assigned to one person and one person only; universal, covering the population involved or at risk; permanently assigned; and reliable. In most systems an assigned code is backed up with other codes such as automobile license number, Social Security number, name codes, and other personal information. The standard identifier in computer usage makes for ease of processing, reduces human intervention and involvement, lessens communications problems, and makes for greater stability in the system. According to reports of the American National Standards Institute, especially of its Task Force on Identification of Individuals and Organizations for Information Exchange, the technical aspects of broad record linkage systems of primary number code and name links have been worked out. An example from the business world is the Data Universal System (DUNS), and examples from government include the Employer Identification Number (EIN) of the Internal Revenue Service, the Social Security Number (SSN), and the Social Insurance Number (SIN) of Canada.[3] It is the purpose of this paper to examine national record linkage systems and in particular those which use a number assigned at birth as the first link in the system.

The Social Security Number

In this country, the Social Security Number is used extensively in identifying and linking records inside and outside the government. The SSN, as an identifier in federal agencies, was authorized by Executive Order in 1943. The Internal Revenue Service, Civil Service Commission, Federal Aviation Agency, Treasury Department, Veterans Administration, and Department of Defense use it as such. The Indian Health Service uses it as an identification number in maintaining comprehensive health histories of Navajo Indians living on federal reservations. In the Veterans Administration, the VA Patient Treatment File basic code number is the SSN; the backup identification is the VA Claim Number. Records summarizing patient episodes are linked together in the system. While it was not originally intended to be used to link records, the SSN is used for this purpose in fact.

Concern with the whole area of record linkage and its implications caused the National Center for Health Statistics to establish a study group on the subject in 1964 in the Public Health Conference on Records and Statistics (PHCRS). At the biennial meeting in 1966, Anita Bahn strongly recommended that the SSN be provided in the Standard Certificate of Birth, stating that "this procedure would complete the individual's vital records linkage and aid immeasurably in health, population, genetic, social and economic research."[4]

Officials of the National Center for Health Statistics and the Social Security Administration met on several occasions to discuss the possible development of a new SSN to be issued at birth but no action toward implementation resulted. In 1971 a Social Security Number Task Force recommended that, as the SSN was in fact becoming increasingly universal in computer data collection and exchange (or linkage) and this brought both benefits and potential dangers to society, a national policy should be established on computer data exchange. The committee recognized that the number had ceased to be a "Social Security Number" and that it was being used extensively outside of the federal government. Yet it recommended that the Social Security Administration should cooperate with public and private organizations using the SSN for health, welfare, and educational purposes, and that the Administration should support the use of SSN as a research tool. It also raised questions as to how universal a number it should become, how early in life it should be issued, and what restrictions should be placed on the number both inside and outside the government.[5]

Subsequently, in 1972, Congress passed an amendment to the Social Security legislation in PL 92-603, which extended the use of the SSN to include the following persons in the numbering system: all aliens, any individual who is an applicant for or a recipient of federal funds, including children, any other individual not previously assigned a number, all children below school age (on request of parents), and children of school age at the time of their enrollment in school (optional).[6] The number therefore could be used for all persons in the population beginning with 5 or 6 years of age, and many others besides.

Subsequently, a Department of Health, Education, and Welfare (DHEW) Report of the Secretary's Advisory Committee on Automated Personal Data Systems stated that the federal government has been in the forefront of expanding the use of the SSN. The committee took the position that a standard universal identifier should not be established in the United States now or in the foreseeable future, and recommended "against the adoption of any nationwide, standard, personal identification format, with or without the SSN, that would enhance the likelihood of arbitrary or uncontrolled linkage of records about people, particularly between government or government supported automated personal data systems."[7]

In its present form the SSN does not have all of the characteristics of a personal identifier. Duplicate numbers have been issued and more than 4 million persons have two or more SSNs. The present system does not cover the whole population and it is not arbitrary. Further the SSN contains no check digits for error correction in the computer. Yet it is possible that, with modifications, the SSN could become unique and universal for this country.

The Birth Number

Another idea for record linkage in the United States that has not been mentioned for several years is the universal birth number, a personal identifier provided on the certificate of live birth. In the 1940s the adoption of such a number almost became a reality as a result of suggestions for the reorganization of the vital statistics registration system. State and local vital statistics offices were swamped with requests for birth certificates when World War II started.

The Association of State and Territorial Health Officers urged the establishment of a Vital Records Commission to investigate the system. The Commission recommended that files of information on individuals be coordinated "by tying one set of records to another through a common factor namely a fixed identity number which would be given each person in a national registration." Following this, a Bureau of the Budget report recommended that "A basis for linkage of records could be produced all at once out of a general registry of identity, or it could grow slowly over a span of years through the assignment at birth of a number which could be carried on subsequently filed official documents relating to each individual."[8]

Halbert L. Dunn, MD, head of the national vital statistics office first in the Census Bureau and later in the Public Health Service, suggested establishing a Life Records Index in the state of birth for the purpose of linking certificates of birth, marriage, divorce, and death by using the birth certificate number as the common identifier. He advocated the use of this number by state and federal agencies, suggesting that "the establishment of a nation-wide system of record linkage for all persons in the country will become an invaluable adjunct to the administration of health and welfare organizations and at the same time produce coordinated statistical knowledge of great value."[9] A Council of Vital Records and Vital Statistics, established by the Census and the American Association of Registration Executives (AARE), suggested in 1947 that all states and independent registration areas standardize the birth certificate so that it could be used throughout the United States; the code even allowed for international usage. It suggested a uniform number of three segments, constructed as follows:

U.S. and Registration Areas	Year of Birth	State Serial No.
1	01–56	00	000000

A person born in New Jersey on January 1, 1946, could be given the number: 131-46-000001. As most registrars supported the concept, the AARE recommended that the numbering system be adopted and that it be put into effect January 1, 1949. Public reaction was immediate; a number of magazine articles criticized the scheme, and the numbering concept seemed to disappear from view. It reappeared in the mid-1960s. In 1966 the PHCRS Study Group on Record Linkage of the PHCRS National Center for Health Statistics recommended as follows: "It was the sense of the Study Group that use of a unique birth number based on an area code . . . be established on a State-by-State basis and designated as a pre-printed number on the new birth certificate which will become effective in 1968 and that the death certificate provide space for this number."[10] In its June, 1966, meeting the American Association for Vital Records and Public Health Statistics resolved that the birth number be placed on all state certificates by January 1, 1968. This was done and since 1968, therefore, a "unique and universal" identifier suitable for record linkage with some modification appears on all birth certificates in this country. No extended use has been made of this number except for its original purpose as a birth certificate index number but it could become a valid national standard identifer by the addition, possibly, of a birth date and check digits.

The Birth Number in Other Countries

The use of a unique number issued at birth to all persons in a national population, and providing for all immigrants, is now fully established in several countries.[11] Although developed for administrative accounting purposes and in many cases as an adjunct to existing population registers, the record linkage potentials of the systems have been recognized and now are being more fully exploited. In every country the use of the systems for combining vital statistics and census data is more or less routine, and today the systems are being expanded to generate more health statistics. In no country is there as yet what may be called a comprehensive computerized health statistics system and statisticians in these countries have few illusions concerning the development of extensive data banks. The process of developing plans for linking records meaningfully takes time and expressed national needs take priority over general research objectives. It is interesting to observe that in these systems where feasible record linkage activities may be designed, there is no move to create an enormous central data bank. What is being done is combining or linking selected items in different subsystems the existence of which are already justified. The unique number makes the job easier, for reasons already stated (Table 1).

In Sweden, a birth numbering system was first established in 1947; it was reorganized by statute in 1967 and 1971. Each person at birth is provided a "person number" for life which consists of 10 digits: date of birth (six), birth serial number (three, with odd for males and even for females), and one check digit. A woman born November 1, 1973, could have the number: 731101-0981. The number is used as the administrative code number in most government offices. A number of special registers derive from it and it is used in the production of official statistics related to income, insurance, labor force, and health. The system is linked to vital records and census statistics. A special register of every person born on the 15th of the month containing detailed social and economic information is used as a sampling frame for special studies.[12]

The number is the basis for a central register of all medical personnel. It is the code for linking information in hospi-

TABLE 1—Items in Identification Numbers in National Systems

Country	Total Digits	Date of Birth			Serial No.	Check Digit	Sex	Place of Birth
		Day	Month	Year				
Denmark	10	2	2	2	3	1		
Finland	10	2	2	2	3	1		
France	15		2	2	3	2	1	5
Germany	12	2	2	2	4	1	1	
Iceland	9	2	2	2	2	1		
Norway	11	2	2	2	3	2		
Sweden	10	2	2	2	3	1		

tal discharge reports. The reports cover diagnosis, operations performed, length of stay, and other more general information. A central cancer registry is now in operation and a central confinements registry has been established. New forms provide detailed information on all deliveries in Sweden, and will be used as a statistical supplement to the birth certificate. All of the records mentioned above are coded by the person number and as the system grows it is expected that the area of linkage will be expanded.

In Norway in 1964 all persons residing in the country were included in a new population register using the birth number as the basic identifier.[13] This is an 11-digit number with six for date of birth, three for a serial number (even for women and odd for men), and two check digits. The number for a female born October 27, 1973, could be: 27.10.73 012 78. Extensive use of the unique number has been made. For family record linkage purposes the identification numbers of mother and father appear on the certificates and other records of their children. Husbands' and wives' numbers are included on the marriage certificates. In the 1970 census the identity number of the family head became the "Family Number." Plans are under way to develop a "National Health Number." This medical and health code number will be a record linking tool with hospitals and physicians reporting on morbidity and health care on all patients in the country on a continuous basis.

In Denmark the birth number is the Central Population Register (CPR) number consisting of 10 digits, six for date of birth, three for serial number, and one check digit which also provides for sex identification. A man born December 25, 1973, could be provided the following: 25.12.73 - 2154. The CPR number, used widely throughout the government for administrative purposes, has now been extended to the hospital sector. Patient records past and present are being put on tape using the birth number as the standard identifier. About half of the hospitals are now covered by this system and it will include all of the hospitals by 1976.

To a lesser extent, the same type of birth number is also in use in Finland and Iceland. Building on an earlier identification system established in 1941 for social security purposes, France developed a system modified for computer application in 1971 and 1972.[14] The most complex number now in use, it combines a birth number with a special geographic code. The 15-digit "national identification number" consists of the following digits: one for sex (1—male and 2—female), two for year of birth, two for month of birth, five for place of birth, three for serial number, and two computer check digits. For a man born April, 1969, in Paris XIV, the number could be: 1.69.04.75.114.925.27. Plans by the Institut National de la Statistique et des Etudes Economiques (INSEE) include the linking of birth and death records, and marriage and divorce records. Marriage, natality, and mortality will be studied by region and by subgroups. There is one major shortcoming to the French system; it is not cleared for migration and as a consequence there may be more persons on file than actually live in France.

A birth number will soon be in use in the Federal Republic of Germany.[15] Known as the PK number (Personenkennzeichen), it is similar to those used in Scandinavia except that the seventh digit of the 12-digit number also designates century of birth and sex. There are six digits for date of birth, one for sex and century, four for serial number, and one check digit. A female born October 30, 1972, would be registered under: 30.10.72 - 423412. The overall system, known as the Population System, may become the most sophisticated in the world. Many record linkages are being planned for interagency administrative reference, and it is expected that, from the first, data will be generated regarding health and welfare and statistics related to the family, but it is too soon to anticipate what form these will take as the legislation regarding this new numbering system is still before the parliament.

In the United Kingdom, an identifier is assigned at birth for use in the National Health Service. It differs from the birth numbers in other countries in several respects. It consists of five letters followed by a birth register entry number; the first three letters indicate place of birth, the fourth indicates year of registration, and the fifth letter the quarter of registration. The number is not used by any other agency; it serves only as an identifier in the documentation of the National Health Service. In recent years the National Health Service Central Index has been used for epidemiological studies and operates as a kind of population register in research programs but considerations of confidentiality have restricted extended use of the system.

In Israel, a population register number is issued to each person born or immigrating, but is unlike the numbers indicated above in that it is in no way linked to the personal characteristics of an individual or a location. This register is updated each month and special tabulations are prepared on various subjects which may link migration, births, marriages, and deaths. However, the use is chiefly administrative and the statistics are used primarily for planning. They are also used as a census check. Future projects include the development of registers for health and hospitalization statistics.[16, 17]

Other countries reported to be considering development of a birth number system are Argentina, Chile, Belgium, the Netherlands, Luxembourg (together as the Benelux countries), Japan, Switzerland, Spain, South Korea, the German Democratic Republic, and Uruguay. As in other countries, initial interest begins with developing a computer accounting system for administrative planning and record keeping and then it moves to many other areas, including health.

Conclusion

It may be only a matter of time before computer accounting and retrieval techniques will combine a large number of records in national health systems. A universal identity number or standard identifier is now being used as a primary link for such systems in various countries, and in other areas there is interest in using such a number. An important next step will be to dispel the public confusion that exists regarding the use of records on individuals for scientific research purposes, on the one hand, and the misuse of records leading to the invasion of privacy, on the other. Precautionary steps have already been taken in those coun-

tries where national systems have been introduced; legislation has been drafted to ensure confidentiality and to limit cross-reference of data except for stated administrative and research purposes with access available only to authorized persons.

There is a significant role for the public health statistician and the researcher in this development. The purpose of record linkage in health and the nature of statistical research should be made widely known. As the country moves in the direction of expanding record linkage systems the statistician should take his place among the advisors and decision-makers, a situation which apparently does not exist at the present time. As regards large scale retrieval systems the 1971 Social Security Task Force has already gone far ahead in raising policy questions for the future which do not delimit the area for discussion but broaden it. The DHEW Secretary's Advisory Committee on Automated Personal Data Systems, while opposing the establishment of a standard universal identifier and an overall record linkage system, has laid a solid ground for argument and further discussion. The public health statistician should become aware of the scope and implications of the discussions as they touch upon his work and the development of the techniques used in his field. There is a danger that ignorance surrounding the subject of data related to the person may affect the gathering of meaningful health data which in turn will have serious implications for health statistics.

Record linkage systems using a personal and unique number provided at birth have been developed. The United States, as it moves into such areas as national health insurance and other programs, may consider similar systems. Such development could provide many opportunities for health statistics research and development.

REFERENCES

1. Newcombe, H. B., James, A. P., and Axford, S. J. Family Linkage of Vital and Health Records. Atomic Energy of Canada, July, 1957, Reprinted August, 1961, and May, 1964.
2. Acheson, E. D. The Oxford Record Linkage Study: A Central File of Morbidity and Mortality Records for a Pilot Population. Br. J. Prev. Soc. Med. 18:8-13, 1967.
3. American National Standards Institute. Proposed American National Standard Structure for Identification of Organizations for Information Interchange. Draft document BSR X 3.35, X 368/190, X 3L8.3/99, June 18, 1973.
4. Bahn, A. K., and Bahn, R. Considerations in Using Social Security Numbers on Birth Certificates for Research Purposes. Public Health Rep. 79:937-938, 1964.
5. Social Security Administration. Social Security Number Task Force: Report to the Commissioner, p. 72. Social Security Administration, Washington, DC, 1971.
6. An Act to Amend the Social Security Act and for Other Purposes. PL92-603, 92nd Congress, HR-1, Section 137, pp. 37-38. Washington, DC, Oct. 30, 1972.
7. U.S. Department of Health, Education, and Welfare. Records, Computers, and the Rights of Citizens. DHEW Publication No. (OS) 73-94, p. 112. DHEW, Washington, DC, 1973.
8. Measures Relating to Vital Records and Vital Statistics. 78th Congress, 1st Session, House Document 242, p. 19. Washington, DC, 1943.
9. Dunn, H. L. Record Linkage. J. A. M. A. 36:1412-1416, 1946.
10. Study Group Reports of the PHCRS. PHCRS Doc. 612.1, p. 66. DHEW, Washington, DC.
11. Lunde, A. S. The Birth Number in Population Statistics. 1973 General Conference of I.U.S.S.P., Liege. No. 1973-743-936/7770, Region 4. U.S. Government Printing Office, Washington, DC, 1973.
12. Description of the Swedish National Registration of Population. Swedish Tax Board Document pp. 10-25, 32-37. Stockholm, 1971.
13. Selmer, E. S. Registration Numbers in Norway: Some Applied Number Theory and Psychology. J. R. Stat. Soc. 130:225-231, 1967.
14. L'Institut National de la Statistique et des Etudes Economiques. Le Numero National d'Identité Description. No. 160/429. Paris, Jan. 19, 1971.
15. Statistisches Bundesamt. Betriftt: Personen Kennzeichen. Federal Republic of Germany, 1972.
16. Bachi, R., Baron, R., and Nathan, G. Methods of Record Linkage and Applications in Israel. Bull. ISI 42:766-782, 1969.
17. Population Register of Israel. State of Israel, Ministry of Interior, Jerusalem, Feb., 1972.

HEALTH CARE PUBLICATION AVAILABLE

The American-Scandinavian Foundation (ASF) announced publication, on September 1, of a special issue of its magazine, *Scandinavian Review*, on health care in Scandinavia. Long sought by American health planners and students of international medicine, this publication is the first authoritative treatment in English of the health care systems in Denmark, Finland, Iceland, Norway, and Sweden. The ASF publication provides a candid comparison between the various Scandinavian approaches to health care delivery and those proposed and already in effect in the U.S.

Single copies of the ASF publication on health care in Scandinavia are $4.95 each. Information on quantity discounts for orders of 25 copies or more is available on request from The American-Scandinavian Foundation, 127 E. 73rd Street, New York, NY 10021.